

Revisão do trabalho de investigação

Transferência de Estilo em Imagens com ViT

¹ Universidade do Minho, Escola de Engenharia

² Mestrado em Engenharia Informática

³ Aprendizagem profunda

Resumo: Nesta análise ao trabalho “*Transferência de Estilo em Imagens com ViT*” será feita uma travessia pelo conteúdo de todas as partes (secções e subsecções) que compõem o trabalho realizado. Serão apontados aspectos positivos e aspectos passíveis de melhoria em cada uma delas. No final, será apresentado um comentário do trabalho na sua generalidade.

1 Análise ao conteúdo do trabalho

1.1 *Abstract*

O *abstract* deste trabalho, de um modo geral, está bem elaborado. É feita uma ténue introdução ao tema, apresenta-se a estrutura do documento do trabalho de forma clara e é feita uma breve menção às conclusões. Este *abstract* transmite com facilidade a estrutura do trabalho, dando ao leitor uma boa visão geral preliminar do seu conteúdo.

1.2 Introdução

A introdução inicia com a explicação do conceito de transferência de estilo. A explicação, apesar de ser curta, está muito bem feita e permitiu entender com bastante facilidade a noção abstracta (isto é, excluindo detalhes técnicos) do conceito. Para além disso, a figura utilizada (fig. 1) complementa muito bem a explicação verbal, fornecendo uma boa noção visual do funcionamento da transferência de estilo.

Relativamente às aplicações (secção 1.1), considero pertinente ter sido mencionada a relevância destas tecnologias em certas indústrias como a moda, arquitectura e arte. No entanto, na parte em que se menciona a empresa *Prisma Labs*, apenas é referido que a mesma usa estas tecnologias para fornecer um certo serviço aos seus clientes. Seria interessante ter sido brevemente explicado no que consiste esse serviço, quais são as suas funcionalidades, etc.

No que diz respeito aos modelos (secção 1.2), esta subsecção fornece uma visão muito superficial de cada modelo, o que é compreensível tendo em conta que os modelos são abordados com mais detalhe na secção do estado da arte. No entanto, esta subsecção não acrescenta muito ao conteúdo já que os modelos são abordados mais à frente. Seria mais enriquecedor se a parte das aplicações fosse estendida (detalhando mais no que estas tecnologias estão a ser utilizadas hoje) e esta parte fosse eventualmente retirada.

1.3 Estado da Arte

1.3.1 *Neural Style Transfer*

Nesta subsecção, a explicação da técnica NST começa com uma visão geral do seu funcionamento, seguida pela descrição de como o NST extrai várias estatísticas das imagens recorrendo às CNNs. Posteriormente, é apresentada uma análise ligeiramente mais detalhada do funcionamento do NST, concluindo com a referência a um desafio desta técnica: encontrar um equilíbrio na importância atribuída ao estilo e ao conteúdo das imagens fornecidas ao modelo.

A partir do que é fornecido nesta subsecção, é possível ter uma compreensão básica (isto é, não muito aprofundada) do NST. No entanto, seria pertinente clarificar os conceitos de estatísticas de conteúdo e estatísticas de estilo (podia ter sido mencionado brevemente o que são essas estatísticas).

Relativamente à limitação de encontrar o equilíbrio entre a ênfase atribuída ao conteúdo e ao estilo, pode-se dizer que esta está bem explicada e é possível entender de imediato qual o problema do NST nesse sentido. Contudo, poderia ter sido abordado com mais detalhe o modo como a função de *loss* faz esse equilíbrio (apenas é dito que existem duas variáveis que fazem esse equilíbrio).

Para finalizar, seria relevante terem sido mencionadas algumas aplicações práticas do NST. Por exemplo, edição de imagem.

1.3.2 *CycleGan*

Esta subsecção aborda a técnica do *CycleGan* que, segundo o que é dito, serve para aplicar o estilo de um ou mais conjuntos de imagens a uma certa imagem de *input*. Para além disso, é bastante reforçado o facto de que esta técnica não necessita de dados que tenham um mapeamento entre o domínio de origem e o domínio de destino.

A explicação fornecida permite obter uma compreensão superficial do *CycleGan* e a figura ilustrativa que foi utilizada complementa bem a explicação fornecida.

A parte final onde são abordados os conceitos de *Adversarial loss* e *Cycle consistency loss*, é possível que não se entenda com clareza esses dois conceitos. Isto acontece porque nesta parte não é explicado no que consistem os mapeamentos F e G . Se esses dois mapeamentos fossem explicados, a compreensão talvez fosse facilitada. Contudo, de acordo com o que fui capaz de inferir, o mapeamento G parece ser a “tradução” da imagem de *input* para o estilo das imagens do domínio alvo e o mapeamento F parece ser o inverso do mapeamento G . Se for esse o caso, então já é possível perceber melhor no que consiste o *Adversarial loss* e o *Cycle consistency loss*.

Para finalizar, pode-se destacar o aspecto positivo de terem apontado alguns casos práticos de aplicação desta técnica.

1.3.3 *Visual Transformers (ViT)*

Esta subsecção inicia com uma exposição da motivação à utilização de *Transformers* (criados para o processamento de linguagem natural), expondo as características promissoras desse método para a área de visão computacional. Não entendo o que se pretende dizer com “... é gratuito aprender a informação global do *input* com a ajuda do mecanismo de *self-attention* ...” (não sei se se pretende indicar que é computacionalmente leve ou se tem outro significado).

Fine-Grained Image Style Transfer with Visual Transformers

A explicação sobre o funcionamento do STTR tem algum nível de detalhe. O processo de funcionamento é descrito passo a passo, desde a conversão das imagens em *tokens* visuais até à obtenção das transformações de estilo. A figura ilustrativa utilizada para complementar a explicação é relativamente complexa, mas permite ter alguma compreensão na arquitectura do STTR.

Apesar de não ser possível obter uma compreensão total do funcionamento desta técnica, o que é compreensível tendo em conta o limite de páginas, foi possível obter uma boa visão geral do funcionamento do STTR. Considero que foi interessante terem fornecido um *link* para o código deste modelo no Github.

Image Style Transfer with Transformers

Tal como no STTR, a explicação do *StyleTr²* apresenta um nível de detalhe satisfatório, possibilitando a obtenção de uma compreensão abrangente do funcionamento desta técnica. A descrição da sua arquitectura também se encontra bem elaborada, contribuindo para a clareza da explicação.

Considero especialmente relevante terem apresentado uma tabela comparativa onde são apresentados resultados de desempenho diversas outras técnicas a competir com esta. Seria interessante ter apresentado algo semelhante nas anteriores, pois daria ao leitor uma noção da qualidade de todas as técnicas. Contudo, entendo que nem sempre possa ser fácil encontrar esse tipo de informações.

1.3.4 Comparação de abordagens

Esta comparação das diferentes técnicas está, de um modo geral, bem feita. Na parte inicial consegue-se perceber bem que os *Transformers* não são tão eficazes com o uso de abordagens tradicionais devido ao seu enviesamento nas formas das imagens e não nas texturas. No entanto, não deu para entender com clareza a parte que se refere ao facto dos *Transformers* apresentarem melhores resultados quando são retreinados com uma supervisão de estilo adequada. Poderiam explicar melhor quando é que é necessário retrainar o *transformer* e no que consiste a supervisão de estilo.

A segunda parte em que se compara o STTR com o StyTR², o foco da comparação é maioritariamente colocado no StyTR². Até pode ser compreensível, uma vez que, pelo que é dito, é a técnica que apresenta ter mais problemas. No fim desta parte é referido que o StyTR² foca-se no nível regional e o STTR foca-se no nível semântico. Isto fica um pouco difícil de compreender, pois os conceitos de nível semântico e regional não são explicados.

No fim, concluem que os modelos com *Transformers* têm um desempenho superior, mas não incluem resultados para demonstrar essa comparação. Se esses resultados são os que foram apresentados na figura 5, então este comentário fica sem efeito.

1.4 Trabalho futuro

Esta secção aborda a planificação que o grupo fez para elaborar o trabalho prático. É relativamente complicado fazer uma análise crítica a esta parte, já que cada grupo tem o seu método de trabalho e não tenho autoridade para o questionar. Desta forma, a minha análise será baseada na forma como a planificação foi exposta neste trabalho.

Relativamente à definição de objectivos, pode-se dizer que estes foram definidos de forma clara. Acho importante terem tomado uma atitude conservadora no que diz respeito às expectativas dos resultados obtidos no trabalho. Também considero que a implementação ser baseada em arquiteturas já pensadas é bom, pois o grupo não iniciará o trabalho “do zero”.

No que diz respeito à calendarização, as tarefas estão bem definidas com as datas devidamente planeadas. Não é possível fazer uma análise crítica a este ponto já que o que está em jogo nesta parte é o planeamento do grupo.

Para finalizar, pode-se dizer que é bom o grupo já ter os *datasets* definidos para implementar o trabalho para não gastar tempo a procurá-los.

1.5 Conclusão

A conclusão, apesar de curta, apresenta o essencial, que são as assunções derivadas do estudo que o grupo fez. Neste caso, foram capazes de concluir que os modelos com *Transformers* têm um melhor desempenho do que as abordagens mais tradicionais, fazendo com que os *Transformers* façam parte do seu planeamento da implementação do projecto prático.

2 Comentário final

Este trabalho tem uma estrutura simples que segue uma ordem lógica e adequada (Introdução → Estado da arte (finalizada com uma comparação das técnicas analisadas) → Conclusões).

No geral, o trabalho está bem feito. A selecção do conteúdo foi bem feita e as técnicas do estado da arte foram relativamente bem explicadas, permitindo uma compreensão geral de cada uma. No entanto,

nalgumas partes do trabalho existe alguma carência de detalhe nalguns conceitos. A planificação do trabalho prático também está bem apresentada em que se pode ver que o grupo já antecipou algumas coisas que lhe permitirá agilizar e terminar com sucesso o projecto.